

## 多変量解析での臆予測の良さ

大阪市大 商 石井 吾郎

### §1. adequate 統計量

この§では§2, 3で必要とする範囲で *adequacy* の話を Bahadur (A. M. S. 1954) - Skibinsky (A. M. S. 1967) - 竹内 (聖済学論叢) - 杉浦 (大阪市大修士論文) - 森本 (1968 数学会) より引用する。

#### Notations

$X \in \mathcal{X}$  : 観測し得る確率変数

$Y \in \mathcal{Y}$  : 未来の確率変数

$Z = (X, Y) \in \mathcal{Z}$  :

$(\mathcal{Z}, \mathcal{A}, \lambda)$  :  $\mathcal{A}$  は  $\mathcal{Z}$  上の  $\sigma$ -field,  $\lambda$  は  $\sigma$ -finite な測度

$\mathcal{P} = \{P_\theta; \theta \in \Theta\}$  :  $\mathcal{Z}$  上の確率分布族,  $P_\theta(z)$  は  $\lambda$  に関する密度関数

$\mathcal{B} \subset \mathcal{A}$  :  $\mathcal{A}$  の部分  $\sigma$ -field で  $X$  により導かれるもの.  $\mathcal{Z}$  上の  $\sigma$ -field を  $\mathcal{F}$

とするとき

$C \subset \mathcal{A}$  :  $\mathcal{A}$  の部分  $\sigma$ -field  $C$  により導かれるもの.  $Y$  上の  $\sigma$ -field を  $\mathcal{G}$  とすると

$$C = \{A; A = Y^{-1}(G) \in \mathcal{A}, G \in \mathcal{G}\}$$

$\mathcal{B}_0 \subset \mathcal{B}$  :  $\mathcal{B}$  の部分  $\sigma$ -field で  $X$  の関数  $T(X)$  により導かれるもの.  $T(X)$  の値域を  $T$ ,  $T$  上の  $\sigma$ -field を  $\mathcal{D}$  とし

$$\mathcal{B}_0 = \{A; A = X^{-1}T^{-1}(D) \in \mathcal{A}, D \in \mathcal{D}\}$$

定義.  $\mathcal{B}_0$  が  $\mathcal{B}$  に対し  $C, \mathcal{P}$  に関して adequate である

$\Leftrightarrow$  ① sufficient

$$\forall B \in \mathcal{B}, \exists \varphi_B(z) (\mathcal{B}_0\text{-m.f.}) \quad \forall \theta \in \Theta \text{ に対し}$$

$$\varphi_B(z) = E_\theta(I_B(z) | \mathcal{B}_0), [\mathcal{A}, \mathcal{P}]$$

② transitive

$$\forall C \in \mathcal{C}, \forall \theta \in \Theta \text{ に対し}$$

$$E_\theta(I_C(z) | \mathcal{B}) = E_\theta(I_C(z) | \mathcal{B}_0), [\mathcal{A}, \mathcal{P}]$$

但し  $I_A(z)$  は  $A$  の指示関数,  $[\mathcal{A}, \mathcal{P}]$  は a.e. の意味.

このとき  $\mathcal{B}_0 \text{ adg}(\mathcal{B}; C, \mathcal{P})$

と書き  $T(X)$  を adequate 統計量と呼ぶ.

このとき次の定理が成り立つ.

I  $B \perp\!\!\!\perp C$  ならば ( $\perp\!\!\!\perp$  は独立の意味)

$$B \circ \text{adg}(B; C, \mathcal{P}) \Leftrightarrow B \circ \text{suff}(B, \mathcal{P})$$

II  $Z = X \otimes Y$ ,  $\mathcal{A} = \{B' \otimes C'\}$ ,  $\mathcal{B} = \{B' \otimes Y\}$

$$C = \{X \otimes C'\}$$

$$\mathcal{P} \ll \lambda = \lambda_1 \otimes \lambda_2$$

なるとき

$$B \circ \text{adg}(B; C, \mathcal{P})$$

$\Leftrightarrow$

$$P_\theta(z) = g(z) h_\theta(z)$$

但し  $g$  は  $\mathcal{B}$  可測,  $h$  は  $\mathcal{B} \vee C$  可測, すなわち

$$P_\theta(z) = g(x) h_\theta(T(x), y)$$

III  $Y$  に関する予測問題においては  $X$  に基づく決定関数のリスクと同じかあるいはそれ以下のリスクを持つ  $T(x)$  の関数である決定関数が存在する。

すなわち  $Y$  に関する予測問題においては adequate 統計量  $T(x)$  の関数になっている決定関数にかぎつてよいわけである。以下では  $T(x)$  が観測されたと考えて  $T(x)$  を単に  $X$  と書く。又  $h_\theta(T(x), y)$  を  $f(x, y; \theta)$  と書く。

定義. 不偏予測量,  $r(x)$  が  $Y$  の不偏予測量である

$$\Leftrightarrow \forall \theta \in \Theta \text{ に対し } E_\theta^{X, Y}(Y - r(x)) = 0$$

## § 2. 点予測における良さ

## Notations

$$X = (X_1, X_2, \dots, X_m) \in \mathcal{X} \subset R^m$$

$$Y = (Y_1, Y_2, \dots, Y_n) \in \mathcal{Y} \subset R^n$$

$$\theta = (\theta_1, \theta_2, \dots, \theta_k) \in \Theta \subset R^k$$

$$f(x, y; \theta) = p(x; \theta) g(y; \theta | x)$$

$$r(x) = (r_1(x), r_2(x), \dots, r_m(x))$$

$$\mathcal{R} = \{r(x); E_{\theta}^{x, y}(Y - r(x)) = 0, \theta \in \Theta\}$$

: 不偏予測量全体

$$E_{\theta}^{y}(Y | X=x) = g(x, \theta) = (g_1(x, \theta), \dots, g_n(x, \theta))$$

$$E_{\theta}^{x}(Y) = E_{\theta}^{x}(g(x, \theta)) = h(\theta) = (h_1(\theta), \dots, h_n(\theta))$$

$$V = V_{\theta}(Y, r(x)) = E_{\theta}^{x, y}(Y - r(x))(Y - r(x))'$$

$$V' = V_{\theta}'(Y) = E_{\theta}^{x, y}(Y - g(x, \theta))(Y - g(x, \theta))'$$

$$V^2 = V_{\theta}^2(r(x)) = E_{\theta}^{x}(g(x, \theta) - r(x))(g(x, \theta) - r(x))'$$

$$\phi_{i_1 \dots i_k} = \phi_{i_1 \dots i_k}(x, \theta) = \begin{cases} \frac{1}{p(x; \theta)} \frac{\partial^{i_1 + \dots + i_k}}{\partial \theta_1^{i_1} \dots \partial \theta_k^{i_k}} p(x; \theta) & \text{if } p > 0 \\ 0 & \text{if } p = 0 \end{cases}$$

$$i_1 + i_2 + \dots + i_k \leq \alpha$$

$$h_{i_1 \dots i_k} = h_{i_1 \dots i_k}(\theta) = \frac{\partial^{i_1 + \dots + i_k}}{\partial \theta_1^{i_1} \dots \partial \theta_k^{i_k}} h(\theta)' \quad (\text{列ベクトル})$$

$$i_1 + i_2 + \dots + i_k \leq \alpha$$

$$\Phi = \Phi(x, \theta) = (\phi_{i_1 \dots i_r}, \phi_{0 i_1 \dots i_r}, \dots, \phi_{0 \dots 0 i_1}, \phi_{2 0 \dots 0}, \dots, \phi_{i_1 \dots i_r})$$

$$i_1 + i_2 + \dots + i_r \leq \alpha, \quad L\text{-ベクトル}$$

$$H = H(\theta) = (h_{i_1 \dots i_r}, h_{0 i_1 \dots i_r}, \dots, h_{0 \dots 0 i_1}, h_{2 0 \dots 0}, \dots, h_{i_1 \dots i_r})$$

$$i_1 + i_2 + \dots + i_r \leq \alpha, \quad n \times L \text{ 行列}$$

$$W = W(\theta) = E_{\theta}^{\otimes} (\Phi(x, \theta)' \Phi(x, \theta)) \quad L \times L \text{ 行列}$$

$$= (E_{\theta}^{\otimes} (\phi_{i_1 \dots i_r}(x, \theta) \phi_{j_1 \dots j_r}(x, \theta)))$$

$$B = B(\theta) = E_{\theta}^{\otimes} (g(x, \theta)' \Phi(x, \theta)) - H(\theta), \quad n \times L \text{ 行列}$$

Lemma 1.  $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$ : sym. semi. pos. def.

$$\Rightarrow M_{11} \geq M_{12} M_{22}^{-1} M_{21}$$

但し  $A \geq B$  は  $A - B$  が semi. pos. def. の意味

Lemma 2  $M_{11} = M_{12} M_{22}^{-1} M_{21}$

$$\Leftrightarrow \text{rank } M = \text{rank } M_{22}$$

Lemma 3 確率変数  $X = (X_1, X_2, \dots, X_m)'$  の共分散行列を  $C$  とする.  $\text{rank } C = p < m$

$$\Rightarrow \exists A: (m-p) \times m \text{ 行列}, \text{rank } A = m-p$$

$$A(X' - E(X)) = 0$$

仮定 1.  $p(x; \theta)$  は  $\alpha$  回連続的に微分可能

仮定 2.  $\theta$  の真の値を  $\theta^0$  とする.  $\theta^0$  の近傍  $U(\theta^0) \subset \Theta$  と積分可能な関数  $G(x)$  が存在して,  $\forall \theta \in U(\theta^0)$

に対して  $|f_{i_1, \dots, i_k}(x, \theta) p(x; \theta)| < \infty$

仮定3.  $W(\theta) > 0$  : (pos. def.)  $\forall \theta \in \mathcal{U}(\theta^0)$

以上の準備の下で次の定理が成立する.

定理1. 仮定1, 2, 3が満たされているとき, すべての

$r(x) \in \mathcal{R}$  に対して

$$(1) \quad \nabla_{\theta}^2 (r(x)) \geq B(\theta) W^{-1}(\theta) B(\theta)', \quad \forall \theta \in \mathcal{U}(\theta^0)$$

$\geq$  は実対称行列全体の空間の中で半正値行列全体を  $C$  とするとき  $C$  は凸錐で,  $C$  による半順序の意味での不等式. (1)の右辺を予測問題における *Battacharya* の限界と呼ぶ.

定理2. 定理1で (1) の等号が  $\theta^0$  で成り立つ

$$\Leftrightarrow g(x, \theta^0)' - r(x)' = B(\theta^0) W^{-1}(\theta^0) \psi(x, \theta^0)'$$

同時分布  $f(x, y; \theta)$  に従う確率変数  $X$  を見て  $Y$  を予測するとき  $X$  の関数  $r(x)$  を作って

$$\nabla(Y, r(x)) = E_{\theta^0}^{xy} (Y - r(x))' (Y - r(x))$$

が上の半順序の意味でなるべく小なる  $r(x)$  を求めることを目標とする.

$$E_{\theta^0}^{xy} (Y - r(x))' (Y - r(x)) = E_{\theta^0}^{xy} (Y - g(x, \theta^0))' (Y - g(x, \theta^0)) + E_{\theta^0}^{xy} (g(x, \theta^0) - r(x))' (g(x, \theta^0) - r(x))$$

$$\nabla_{\theta} (Y, r(x)) = \nabla_{\theta} (Y) + \nabla_{\theta}^2 (r(x))$$

この式で  $\nabla^2$  はきまった行列で動かしようのないものである。  
 吾々の出来ることは  $r(x)$  を色々に変えることにより  $\nabla^2$  を小に  
 することである。その  $\nabla^2$  を定理1で評価する。

定理1の証明

$T(x) = (g(x, \theta) - r(x), \Phi(x, \theta))$  とおく,  $n+L$  ベクトル  
 $E_{\theta}^{\otimes} T(x) = 0$  である.  $\because$  仮定より  $E_{\theta}^{\otimes} (g(x, \theta) - r(x)) = 0$

$\int_{\mathcal{X}} p(x; \theta) dx = 1$  と仮定2を用いて

$$\int_{\mathcal{X}} \frac{\partial^{c_1 + \dots + c_k}}{\partial \theta_1^{c_1} \dots \partial \theta_k^{c_k}} p(x; \theta) dx = \int_{\mathcal{X}} \phi_{c_1, \dots, c_k}(x, \theta) p(x; \theta) dx = 0$$

$U(\theta) = E_{\theta}^{\otimes} (T(x)' T(x))$  とおく

$$= \begin{bmatrix} E_{\theta}^{\otimes} (g(x, \theta) - r(x))' (g(x, \theta) - r(x)), E_{\theta}^{\otimes} (g - r)' \Phi \\ E_{\theta}^{\otimes} \Phi(x, \theta)' (g(x, \theta) - r(x)), E_{\theta}^{\otimes} \Phi' \Phi \end{bmatrix} \begin{matrix} n \\ L \end{matrix}$$

$$= \begin{bmatrix} \nabla^2 & \bar{B} \\ \bar{B}' & W \end{bmatrix} \geq 0$$

Lemma 1 より

$$\nabla^2 \geq \bar{B} W^{-1} \bar{B}'$$

$$\int r(x)' p(x; \theta) dx = h(\theta)' \quad (\text{列ベクトル})$$

の両辺を  $\theta$  につき微分することにより

$$E_{\theta}^{\otimes} r(x)' \phi_{c_1, \dots, c_k}(x, \theta) = h_{c_1, \dots, c_k}(\theta)$$

$$E_{\theta}^{\otimes} r(x)' \Phi(x, \theta) = H(\theta)$$

以上により  $\bar{B} = B$  . よって証明された.

定理2の証明.

定理1で等号が成立していると Lemma 2より

$$\text{rank } \bar{U} = \text{rank } \bar{W} = L, \quad \text{Lemma 3より}$$

$\exists A: n \times (n+L)$  行列,  $\text{rank } A = n$

$$A \begin{bmatrix} g(x, \theta)' - r(x)' \\ \Phi(x, \theta)' \end{bmatrix} = 0 \quad (\because ET(x) = 0)$$

$$A = \begin{bmatrix} A_1 & A_2 \\ \underbrace{\quad}_n & \underbrace{\quad}_L \end{bmatrix} \quad \text{と分けて}$$

$$A_1 [g(x, \theta)' - r(x)'] + A_2 \Phi(x, \theta)' = 0$$

$$A_1 E_{\theta}^{\otimes} (g(x, \theta)' - r(x)') \Phi(x, \theta) + A_2 E_{\theta}^{\otimes} \Phi(x, \theta)' \Phi(x, \theta) = 0$$

$$A_1 B + A_2 \bar{W} = 0, \quad A_2 = -A_1 B \bar{W}^{-1}$$

$$A_1 (g(x, \theta)' - r(x)' - B \bar{W}^{-1} \Phi(x, \theta)') = 0$$

ここで  $\text{rank } A_1 = n$  である.  $\therefore A = A_1 [I, -B \bar{W}^{-1}]$

$$n = \text{rank } A \leq \text{rank } A_1 \leq n$$

$$\therefore g(x, \theta)' - r(x)' = B \bar{W}^{-1} \Phi(x, \theta)'$$

逆が成立することは明らか.

定理3  $\bar{X} = [X^1, X^2]$  を adequate 統計量であるとする.

$E_{\theta}^y(Y | \bar{X}) = g(\bar{X}, \theta)$  が  $X^1$  のみの関数であるとする, ( $G(X^1,$

$\theta)$  とおく).  $X^1$  を与えたときの  $X^2$  の条件付分布が各  $X^1$  に対

して Complete であり且つ  $r(\bar{X})$  は  $G(X^1, \theta)$  の不偏推定量

であるならば  $r(\bar{X})$  は  $Y$  の予測量として許容的である.

証明  $f(\mathbf{X}, \theta) = f_1(x^1, \theta) f_2(x^2, \theta(x^1))$  とする.  $f_2(\cdot | x^1)$   
 に従う  $X^2$  を観測して  $G(x^1, \theta)$  の点推定する問題を考えると  $f_2$   
 の complete 性よりそれは許客的である (Lehmann-Scheffé)  

$$V^2 = E_{\theta}^{X^1} E_{\theta}^{X^2} [ [r(x^1, X^2) - G(x^1, \theta)] [r(x^1, X^2) - G(x^1, \theta)] | x^1 ]$$
  
 より  $r(\mathbf{X})$  が  $Y$  の予測量として許客的である.

### § 3. 正規分布に関する点予測

例 1.  $X_i, i=1, 2, \dots, m$  :  $N(\mu, \sigma^2)$  に従う独立な  $m$  個  
 の観測し得る確率変数

$X_i, i=m+1, \dots, m+n$  : 未来の確率変数

とするとき  $X_i, i=1, 2, \dots, m$  を得て

$$Y_1 = \frac{1}{m+n} \sum_{i=1}^{m+n} X_i$$

$$Y_2 = \frac{1}{m+n-1} \sum_{i=1}^{m+n} (X_i - Y_1)^2$$

の点予測を行う.

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad S = \sum_{i=1}^m (X_i - \bar{X})^2$$

$$\tilde{X} = \frac{1}{n} \sum_{i=m+1}^{m+n} X_i, \quad \tilde{S} = \sum_{i=m+1}^{m+n} (X_i - \tilde{X})^2$$

とおく.

adequate 統計量  $\mathbf{X} = (\bar{X}, S)$  ( $S$  の  $I$  より)

$$g_1(\mathbf{X}, \theta) = E(Y_1 | \mathbf{X}) = \frac{1}{m+n} (m\bar{X} + n\mu)$$

$$g_2(\mathbf{X}, \theta) = E(Y_2 | \mathbf{X}) = \frac{1}{m+n-1} \left( S + (n-1)\sigma^2 + \frac{m \cdot n}{m+n} \left( (\bar{x} - \mu)^2 + \frac{\sigma^2}{n} \right) \right)$$

$$r_1(\mathbf{X}) = \bar{x}, \quad r_2(\mathbf{X}) = \frac{1}{m-1} S \quad \text{を } Y_1, Y_2 \text{ の予測量と}$$

する。

$$\nabla = \begin{bmatrix} \frac{n}{(m+n)m} \sigma^2 & 0 \\ 0 & \frac{2n}{(m+n-1)^2(m-1)} \sigma^4 \end{bmatrix} = \nabla^1 + \nabla^2$$

$$\nabla^1 = \begin{bmatrix} \frac{n}{(m+n)^2} \sigma^2 & 0 \\ 0 & \frac{2n((m+n)^2 - n)}{(m+n-1)^2(m+n)^2} \sigma^4 \end{bmatrix}, \quad \nabla^2 = \begin{bmatrix} \frac{n^2 \sigma^2}{(m+n)^2 m} & 0 \\ 0 & \frac{2n^2((m+n)^2 + m-1) \sigma^4}{(m+n-1)^2(m-1)(m+n)^2} \end{bmatrix}$$

$\Phi(\mathbf{X}, \theta) = (\phi_{10}, \phi_{01}, \phi_{20})$  を考えることにより

$$W = \begin{bmatrix} \frac{m}{\sigma^2} & 0 & 0 \\ 0 & \frac{m}{2\sigma^4} & \frac{m}{\sigma^4} \\ 0 & \frac{m}{\sigma^4} & \frac{2m^2}{\sigma^4} \end{bmatrix}, \quad W^{-1} = \begin{bmatrix} \frac{\sigma^2}{m} & 0 & 0 \\ 0 & \frac{2\sigma^4}{m-1} & -\frac{\sigma^4}{m(m-1)} \\ 0 & -\frac{\sigma^4}{m(m-1)} & \frac{\sigma^4}{2m(m-1)} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{-n}{m+n} & 0 & 0 \\ 0 & \frac{-n}{m+n} & \frac{2mn}{(m+n-1)(m+n)} \end{bmatrix}$$

より  $\nabla^2 = B W^{-1} B'$  が成立することかわかる。又

$$g(\mathbf{X}, \theta)' - r(\mathbf{X})' = \begin{bmatrix} \frac{n}{m+n} (\mu - \bar{x}) \\ \frac{1}{m+n-1} \left( \frac{m \cdot n}{m+n} (\bar{x} - \mu)^2 + (n-1 + \frac{m}{m+n}) \sigma^2 - \frac{n}{m-1} S \right) \end{bmatrix}$$

は  $B W^{-1} \Phi'$  に一致する。

例2.  $N(\mu, \sigma)$  での点予測

$$X = (x_1, x_2, \dots, x_p, x_{p+1}, \dots, x_{p+g}) \in R^{p+g}$$

$$= (X^1, X^2)$$

$$X^1 = (x_1, x_2, \dots, x_p), \quad X^2 = (x_{p+1}, \dots, x_{p+g})$$

$$X_i = (x_{i1}, x_{i2}, \dots, x_{ip}, x_{i,p+1}, \dots, x_{i,p+g}), \quad i=1, 2, \dots, m$$

$X_{m+1}^1 = (x_{m+1,1}, \dots, x_{m+1,p})$  を  $X_{m+1}$  の初めの  $p$  個の成分とする。  $X_i, i=1, 2, \dots, m$  と  $X_{m+1}^1$  を観測して  $Y = X_{m+1}^2$  を予測する問題を考える。このときの adequate 統計量は

$$\underline{X} = (\bar{X}, S, X_{m+1}^1)$$

である。但し  $\bar{X}, S$  は  $X_i, i=1, 2, \dots, m$  より作った平均及び  $S = \sum X_i' X_i - m \bar{X}' \bar{X}$  である。

$$r(\underline{X}) = \bar{X}^2 + (X_{m+1}^1 - \bar{X}') S_{11}^{-1} S_{12}$$

は不偏予測量であり、それが最適であることは定理3より示される。